

Cot in the Act: Speaker Ethnicity Conditions Lexical Identification in the Context of the Low-Back Merger in New York City English

Omar Ortiz and Bill Haddican^{1*}

Introduction

Several independent studies have recently reported evidence of THOUGHT-lowering and/or LOT/THOUGHT merging in New York City English led by younger non-White speakers. Spoken corpus data by Wong (2012), Becker (2010) and Haddican et al. (2021) suggest rapid THOUGHT lowering, particularly in Asian and Latinx communities. Similarly, younger Asian and Latinx NYCE speakers have been reported to favor merged LOT/THOUGHT responses in controlled homophony judgment tasks (Johnson 2010, Haddican et al. 2016). Matched-guise results by Becker (2014), moreover, suggest that raised THOUGHT is associated mainly with older White speakers among NY-native judges.

Unaddressed in this literature is whether listeners use perceived social information about the speaker—i.e. perceptions of age and ethnicity—in their phonemic categorization of low back vowels in comprehension of NYCE (Rubin 1992, Hay et al. 2006). Here, we report results from a forced-choice lexical identification experiment intended to investigate this. Consistent with previous production and matched guise results, judges misidentified LOT auditory stimulus items as THOUGHT more often when the item was accompanied by a photo of an Asian speaker than a White speaker. The results suggest that NYCE-native listeners actively use social information about speakers in the categorization of LOT/THOUGHT items in comprehension.

Our discussion is organized as follows. Section two of this paper briefly reviews previous results on THOUGHT-lowering and/or LOT/THOUGHT merging in New York City English. Section three describes the current data set. Section four discusses the results.

THOUGHT lowering in NYCE

A traditional feature of NYCE vowel phonology—one found in descriptions dating to the late nineteenth century—is a distinction between THOUGHT ([ɔ]) and LOT ([ɑ]) lexical sets (Babbitt 1896, Labov 1966, Becker 2010, 2014, Wong 2012, 2014, Labov, Ash, & Boberg 2006, Newman 2014). In addition, NYCE shares with other mid-Atlantic varieties as far south as Baltimore the property of having, among conservative speakers, a sharply raised and slightly ingliding variant [ʊə] for THOUGHT (Labov et al. 2006, Newman 2014).

Over the past decade, however, several independent production studies have reported evidence of lowering of THOUGHT in apparent time in NYCE (Wong 2007, 2010, 2012, Becker 2010, 2014, 2015, Newman 2014, Wong and Hall-Lew 2014, Newlin-Lukowicz 2015, 2016, Haddican et al. 2021). We illustrate this change in the following two plots with data from CoNYCE, a recently developed parsed and audio-aligned corpus of conversational data from contemporary New York City English (Tortora et al., in progress). Figure 1 shows F1 and F2 values for LOT and THOUGHT items illustrating a typical conservative low back pattern. Here, LOT shows fairly little overlap with THOUGHT, which is raised and retracted—some tokens sharply so. Figure 2 illustrates an innovative pattern from a younger Asian-American woman. In this plot, THOUGHT is lower and somewhat fronted, overlapping largely with LOT in acoustic space.

* This work is supported by NSF Grant Numbers BCS-1152148, BCS-1630274, BCS-1630377, BCS-1630286, BCS-1629348, and BCS-1152148.

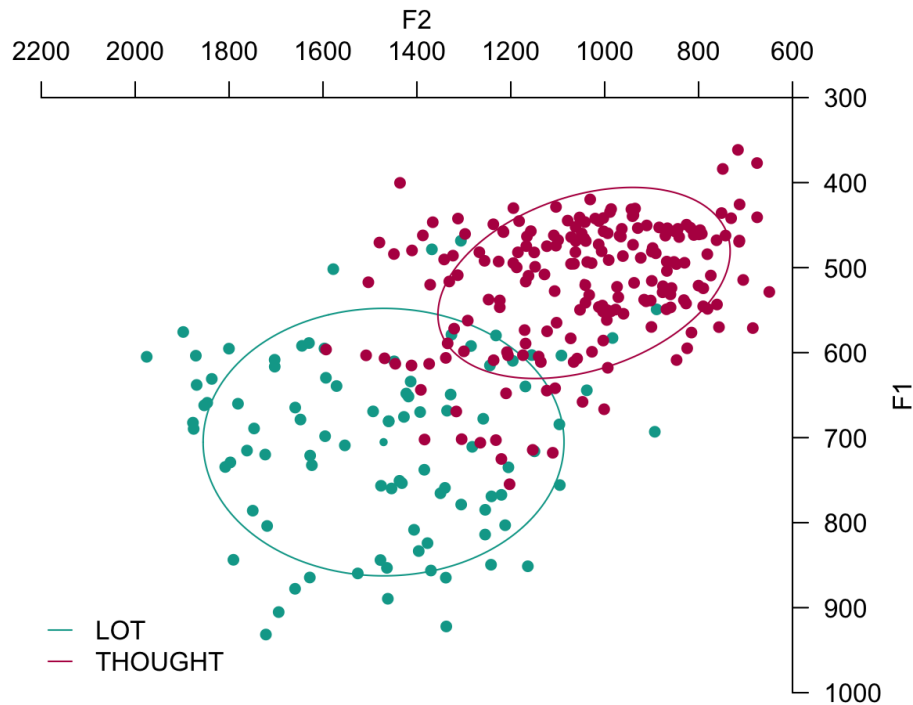


Figure 1: Formant values and confidence ellipses (± 1 standard deviation) for LOT and THOUGHT items from a male White speaker, b. 1959.

Several of the above-mentioned studies suggesting change in low back vowel system have also reported effects of speaker ethnicity on this process of change. In particular, Becker's (2006) data from the Lower East Side of Manhattan shows lowering across ethnicity groups with the exception of African American participants who showed no apparent time evidence of THOUGHT lowering. Haddican et al. (2021) report that raised THOUGHT is best maintained among White speakers across age groups and that the innovative lower realizations for THOUGHT are found largely among non-White New Yorkers, particularly Asian and Latinx speakers. Lowered THOUGHT among Asian speakers is also described in Newman (2014), Wong (2007, 2010, 2012) and Wong and Hall-Lew (2014). Wong and Hall-Lew (2014) suggest that lowered THOUGHT among Asian speakers is related to a strong association between raised THOUGHT and traditional White working-class New York City personas. Evidence for such an indexical link between raised THOUGHT and traditional White personas comes from Becker's (2014) matched guise results, where judges associated raised THOUGHT with a "classic" older White New Yorker persona from the outer boroughs.

While no reports have yet described broad prevalence of LOT-THOUGHT mergers in production, homophony judgement data by Johnson (2007) and Haddican et al. (2016) support an expansion of the merger in perception, particularly among Asian and Latinx New Yorkers. (See also Newman (2014) who observes some incidence of merger among second generation South Asian New Yorkers.) Figure 3 shows participant age results from Haddican et al. (2016)'s homophony judgment study, where participants indicated whether a series of LOT-THOUGHT minimal pairs were pronounced the "same" or "different" in their English. Older speakers tended toward unmerged responses much more than younger speakers. In addition, Haddican et al. (2016) report that merged responses were favored by Latinx, East Asian and South Asian participants relative to Black and White participants, as shown in Figure 4.

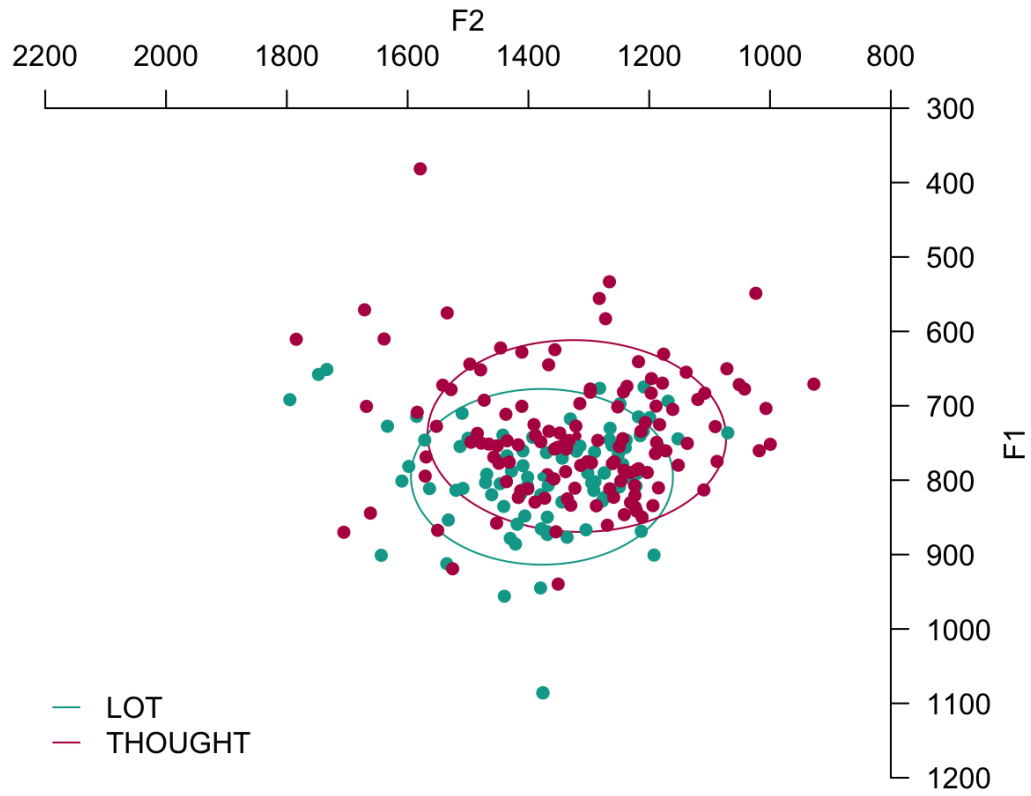


Figure 2: Formant values and confidence ellipses (± 1 standard deviation) for LOT and THOUGHT items from a female Asian-American speaker, b. 1999.

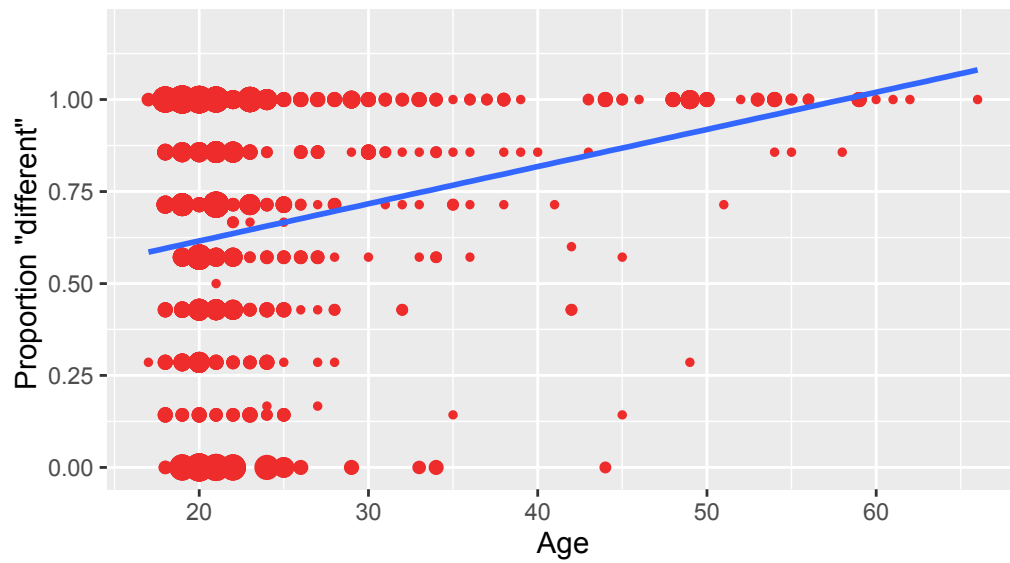


Figure 3: Proportion of merged judgments by age (Haddican et al. 2016).

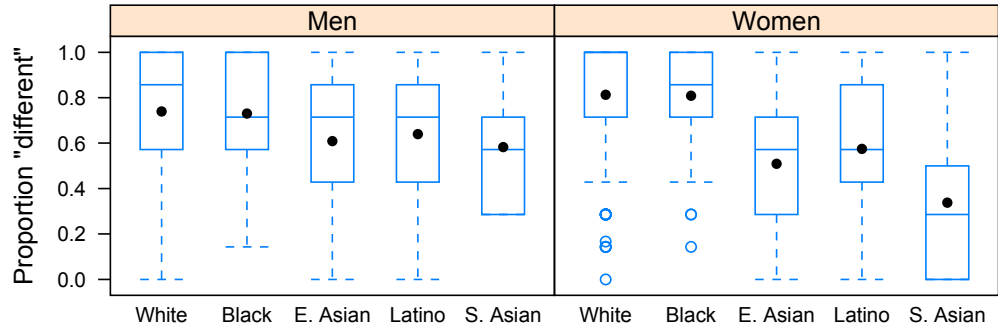


Figure 4: Merged judgments by gender & ethnicity (Haddican et al. 2016).

Unaddressed in previous literature is whether listeners use perceived information about speaker age and ethnicity in their phonemic categorization of low back vowels in comprehension of New York City English in a way similar to that reported by Hay et al. for New Zealand English. In the remaining discussion, we describe a study aimed to address this.

Methods

Much previous work has shown that non-linguistic information conditions listener performance on different speech perception tasks (Rubin 1992, Niedzielski 1999; Hay et al. 2006, Hay & Drager 2010, Mack & Munson 2012). One specific line of work in this area suggests that visual cues to the social categorization of the speaker affect comprehension and/or evaluation of a speaker's production (Rubin 1992, Strand 1999, 2000, Staum 2008, Koops 2011, Babel and Russell 2015, D'Onofrio 2015, 2020, McGowan 2015, Ortiz 2018). Our research design is based on that in Hay et al. (2006), who looked at lexical identification in NEAR-SQUARE minimal pairs—a merger in progress in New Zealand English. Their visual stimuli biased age and class as social categories hypothesized to condition lexical identification for items in ten minimal pairs: *ear/air*, *beer/bare* etc. Participants chose between two items in these pairs given both auditory and visual stimuli. Here, we employ a similarly structured design to test for age and ethnicity effects on lexical identification of LOT-THOUGHT minimal pairs.

Participants were recruited through the Queens College (CUNY) Psychology Subject Pool in the spring of 2020, and received course credit for participating. We recruited 83 self-reported native speakers of New York City English from a range of self-reported language backgrounds, and whose distribution by self-reported gender, ethnicity and age is as follows:

- **Gender:** 62 women, 21 men.
- **Ethnicity:** 37 Asian, 23 White, 23 Other.
- **Median year of birth:** 2001.

Participants completed the experiment online in the spring of 2020 via a Qualtrics survey. In each trial, participants were presented with two stimuli: (i) a picture of a speaker; and (ii) a single-word auditory stimulus from one of six LOT-THOUGHT minimal pairs: *caught-cot*, *taught-tot*, *dawn-Don*, *bought-bot*, *talk-tock*, *auto-Otto* taken from Johnson's (2010) homophony judgment design. Candidate photographs to serve as visual stimuli were taken from free-use online sources. In a prior norming phase, judges (N=75) categorized stimulus photographs of women into one of two age categories (older, $M = 46.5$ years old, $SD = 2.1$ vs. younger $M = 23.4$ years old, $SD = 1.7$)

and one of three ethnic categories (Latina, White, Asian). Photographs selected as stimuli were those with over 88% agreement between norming participants and our intended bias for ethnicity and age categorization.

The stimulus voices were from six self-identified White women and native NYCE speakers whose ages ranged from twenty-eight years old to fifty years old. The six stimulus voices were all LOT-THOUGHT distinguishers, albeit with some variation in their degree of differentiation of these sets in acoustic space. Figure 6 shows raw formant values for all stimulus items by speaker.

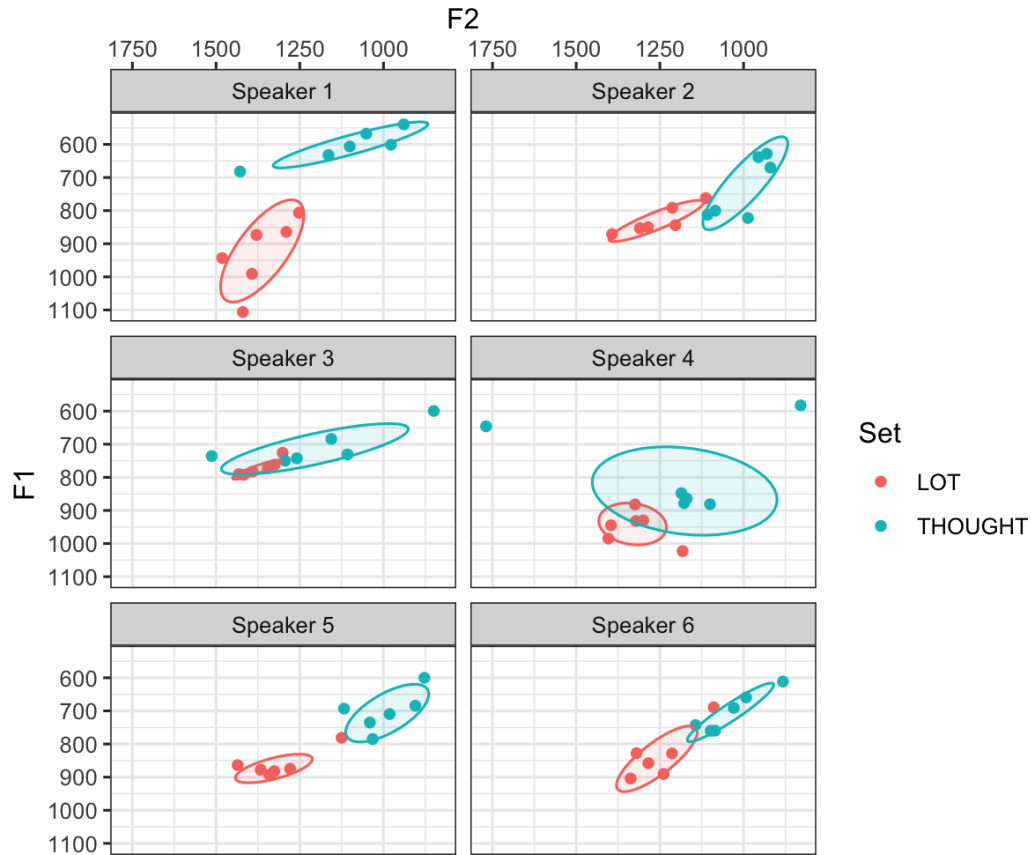


Figure 6: Formant values and confidence ellipses (± 1 standard deviation) for LOT and THOUGHT auditory stimuli.

The stimulus voices and photographs were rotated across six lists, to which subjects were assigned by a counter. In each trial, participants identified a lexical item (using standard English orthography) associated with the auditory stimulus, i.e. [kat] as “caught” or “cot”. Figure 7 is a screenshot of an example trial, with a photograph of a young Asian woman and auditory stimulus [aro]. Participants were asked to press play on the audio clip, and then select which word they heard by clicking one of two buttons whose linear order was randomized across trials. The experimental items consisted of 72 such trials (12 photographs x 6 voices/minimal pairs).

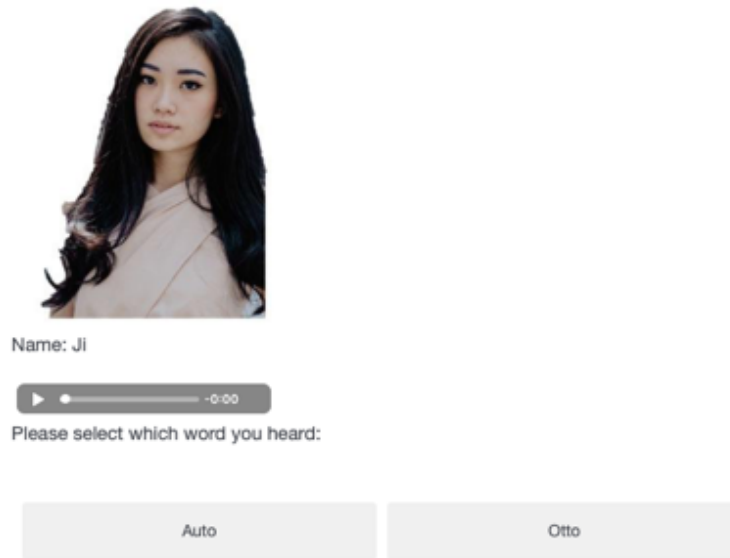


Figure 7: Example trial.

Results and Discussion

The dependent measure used in modeling was accuracy on LOT stimuli, i.e. whether judges correctly identified LOT items as belonging to this class rather than THOUGHT. THOUGHT items were not included in the model since all judges were at or near ceiling for these items, consistent with previous descriptions of low back change in NYCE. That is, because the merger in NYCE is driven principally by THOUGHT lowering rather than LOT raising/backing (Haddican et al. 2021), raised THOUGHT items were likely to be unambiguous for participant judges.

Overall LOT accuracy was above chance with the notable exception of *auto/Otto*, for which LOT accuracy was low, as shown in Figure 8. The exceptional behavior of *auto/Otto* may reflect variable categorization of *auto* in the community, as also suggested by Haddican et al.’s (2016) homophony judgment results for this pair. (See, similarly, Hay et al. (2006) who find that NEAR/SQUARE lexical identification error rates are related to the proximity of items in acoustic space in production.)

Table 1 summarizes a generalized linear mixed effects regression model using the *lme4* package in R (Bates et al. 2015, R Core Team 2021). The analysis reveals two main effects—one for stimulus photo, and a second for participant ethnicity. Importantly, the analysis revealed no effect of stimulus photo age group on LOT accuracy, which we return to shortly.

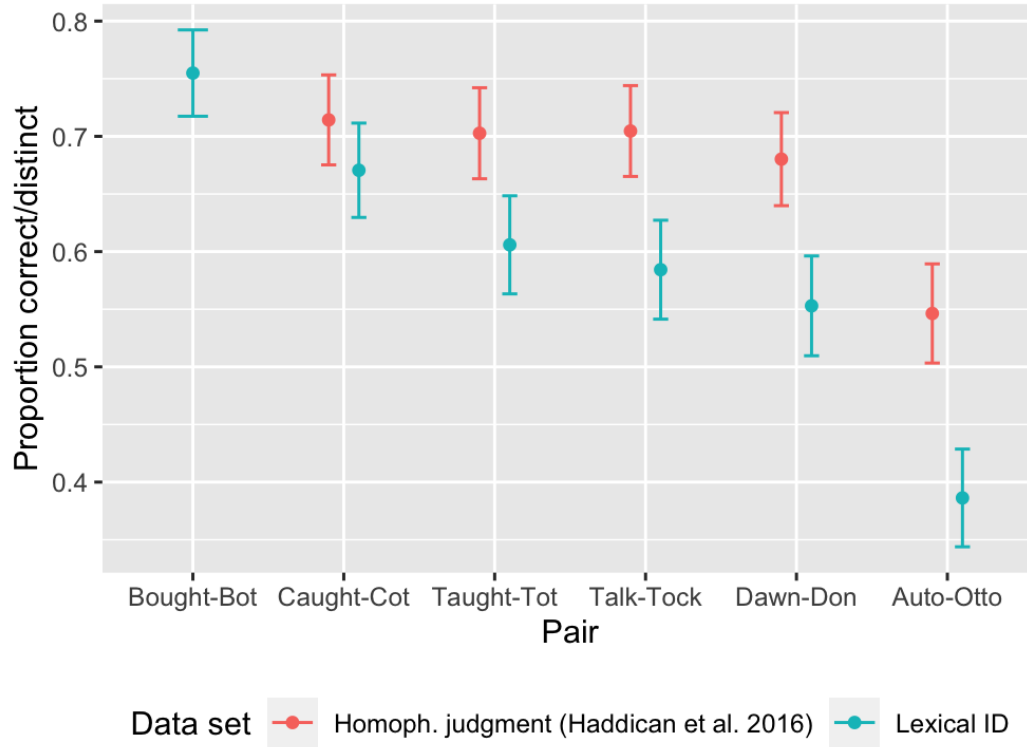


Figure 8: Lexical Identification accuracy and homophony judgments for five pairs. (The *bought/bot* pair was not included in Haddican et al.’s survey.)

Fixed Effects	Estimate	Std. error	z	p
Intercept	0.098	0.257	0.381	0.704
Photo: Latina	0.077	0.098	0.787	0.431
White	0.273	0.099	2.770	0.006
Participant: Other	0.305	0.203	1.499	0.134
White	0.522	0.204	2.559	0.011

Table 1: Summary of GLMER model of correct identification of LOT stimuli. Model formula: Response ~ Stimulus ethnicity + Participant ethnicity + (1 | Participant) + (1 | Item). Obs=2988, N=83. Reference level is “Asian” for both Stimulus Photo and Participant ethnicity. A random slope for photo ethnicity was not fittable for either items or participants.

Figure 9 shows model estimates of LOT item accuracy by stimulus photo. Post-hoc pairwise comparisons using the *emmeans* package in R (Russell 2021) reveals a significant contrast between Asian and White conditions ($p=.016$) but not for Asian/Latina or Latina/White comparisons.² In particular, these results suggest that LOT stimuli were more likely to be misinterpreted as THOUGHT items—i.e. *caught* instead of *cot*—when the speaker photo was of an East Asian face than a White face. This effect of photo ethnicity is therefore consistent with the converging evidence summarized above—(i) production results (Wong 2007, Becker 2010), (ii) homophony judgement surveys (Johnson 2010, Haddican et al. 2016), and (iii) matched guise results (Becker 2014)—all suggesting an effect of ethnicity on THOUGHT lowering and merger of

² P -values were adjusted using the Tukey method.

LOT and THOUGHT. In the present case, these results suggest that listeners use perceived information about speaker ethnicity (White vs. Asian) in lexical identification of LOT-THOUGHT minimal pairs.³

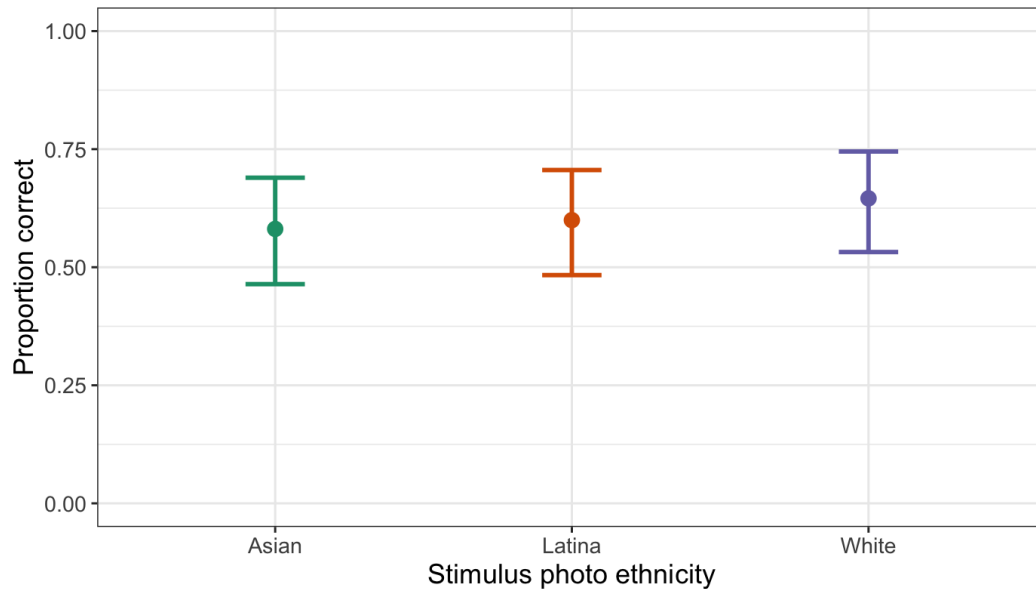


Figure 9: Model estimates for LOT accuracy by stimulus photo condition.

Figure 10 shows model estimates of LOT item accuracy by participant ethnicity group. Here, White participants are more likely to correctly identify LOT items than Asian participants. Again, pairwise comparisons reveal a contrast between White and Asian groups ($p=.028$), but not for Asian/Other and Other/White contrasts. One possible interpretation of this result is that White subjects' higher aggregate accuracy was facilitated by the fact that all of the stimulus speakers were themselves White, that is, that our White judges were more familiar with the vowel systems of White New York speakers.

Another possibility is that these differences reflect properties of the subjects' own low back systems, and in particular whether they themselves have a merger (Hay et al. 2006). That is, if, as discussed above, low THOUGHT and/or the low back merger is more prevalent in Asian than in White communities, then the contrast in Figure 10 could instead reflect a greater prevalence of merger or near-merger in the Asian participant category. We addressed this alternative possibility by including, as a predictor in a separate model, participants' performance on a three-item homophony judgment task included in the instrument. Participants were asked whether minimal pairs *caught-cot*, *dawn-Don*, *taught-tot* "sounded the same" in the context of a carrier sentence, using Johnson's (2010) questionnaire items. Importantly, in this model, distinguishers did not perform better than non-distinguishers, offering no support for the possibility that the between-participant differences reflected in Figure 10 reflect whether participants themselves have a merger in homophony judgments.

³ We have no account for the absence of an effect for the Latina condition. We speculate it may be related to variability across Latinx communities in participation in THOUGHT-lowering, some discussion of which can be found in Newman (2014).

COT IN THE ACT

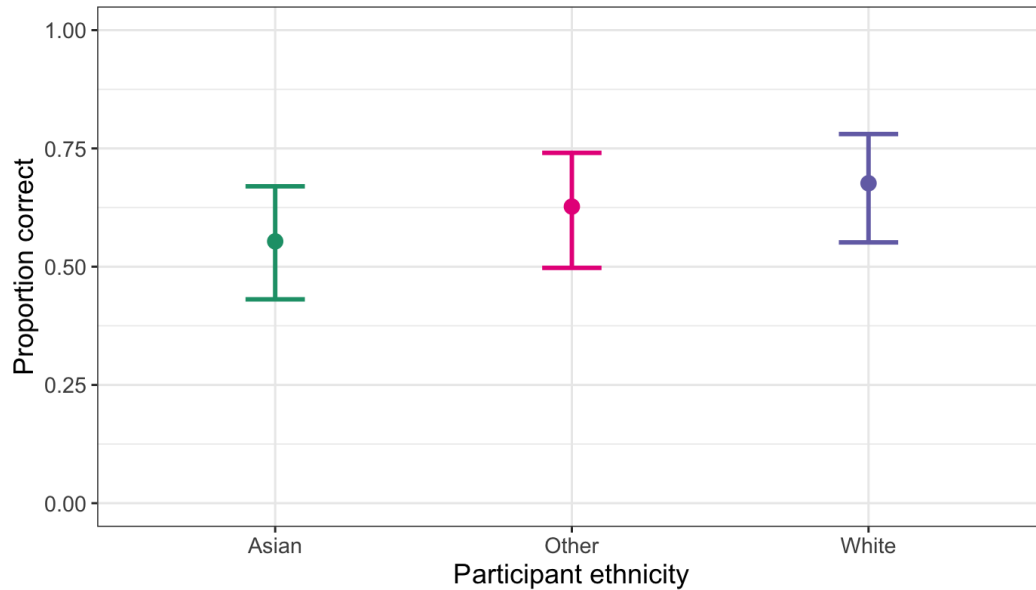


Figure 10: Model estimates for LOT accuracy by participant ethnicity.

Notably absent in our results is an effect of stimulus age. Again, age effects reported in production and homophony judgment studies indicate THOUGHT lowering in progress and incipient LOT-THOUGHT merger, leading to an expectation of a speaker age effect on LOT identification, contrary to results. Possibly relevant to this outcome is that THOUGHT-lowering is not a new change in New York City English. Independent samples by Becker (2010) and Haddican et al. (2021) both suggest apparent time change beginning with speakers born in the first half of the twentieth century. We speculate that the age bias induced by the stimuli ($M=46.5$; $M=23.4$ respectively in the norming data) was insufficient to elicit an effect on lexical identification, given participant knowledge of the range of age effects on THOUGHT lowering in the community.

Conclusion

The present results suggest that listener knowledge of ethnic correlates of THOUGHT lowering is active in listeners' identification of lexical items in New York City English. More generally, this work contributes to a growing body of work suggesting that perceived social information about speakers is recruited in speech perception (Hay et al. 2006, Strand 2000, Staum 2008, Koops 2011, Babel and Russell 2015, D'Onofrio 2015, 2020, Ortiz 2018).

References

- Babbitt, Eugene Howard. 1896. The language of the lower classes in New York City and vicinity. *Dialect Notes* 1:457-464.
- Babel, Molly, and Jamie Russell. 2015. Expectations and speech intelligibility. *The Journal of the Acoustical Society of America* 137(5):2823-2833.
- Bates Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1):1-48.
- Becker, Kara. 2010. Regional dialect features of the Lower East Side of New York City: Sociophonetics, ethnicity, and identity. Doctoral Dissertation, New York University.
- Becker, Kara. 2014. The social motivations of reversal: Raised BOUGHT in New York City English. *Language in Society* 43:395-420.
- Becker, Kara, and Amy Wing-mei Wong. 2010. The short-a system of New York City English: An Update. In *U. Penn Working Papers in Linguistics 15.2: Selected Papers from NWAV 37*, ed. K. Gorman and L. MacKenzie, 11-20.

- D'Onofrio, Annette. 2015. Perceiving Personae: Effects of Social Information on perceptions of TRAP-backing. In *U. Penn Working Papers in Linguistics 21.2: Selected Papers from NWAV 43*, ed. S. Fisher, 31-39.
- D'Onofrio, Annette. 2020. Personae in sociolinguistic variation. *Wiley Interdisciplinary Reviews: Cognitive Science* 11:e1543.
- Haddican, Bill, Daniel Ezra Johnson, Michael Newman, and Faith Kim. 2016. The diffusion of the low back merger in New York City. Poster presented at NWAV 45.
- Haddican, Bill, Cecelia Cutler, Michael Newman, and Christina Tortora. 2021. Cross-speaker covariation across six vocalic changes in New York City English. Ms. CUNY.
- Hay, Jennifer and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4): 865–892.
- Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4):458-484.
- Johnson, Ezra. 2010. *Stability and change along a dialectal boundary: The low vowels of Southeastern New England*. Durham: Duke University Press.
- Koops, Christian. 2011. Local Sociophonetic Knowledge in Speech Perception. Doctoral Dissertation, Rice University.
- Labov, William. 1966. *The social stratification of English in New York City*. 2nd edition. Washington DC: Center for Applied Linguistics.
- Labov, William, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English*. Berlin: Mouton de Gruyter.
- Lenth, Russell V. 2021. Emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.6.3.
- Mack, Sara, & Benjamin Munson. 2012. The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, 40(1): 198-212
- McGowan Kevin B. 2015. *Social Expectation Improves Speech Perception in Noise*. *Language and Speech* 58(4):502-521.
- Newlin-Lukowicz, L., 2016. Co-occurrence of sociolinguistic variables and the construction of ethnic identities. *Lingua* 172:100–115.
- Newman, Michael. 2014. *New York City English*. Berlin: Mouton de Gruyter.
- Niedzielski Nancy. 1999. The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology* 18(1):62-85.
- Ortiz, Omar. 2018. Beyond macro-level ethnicity: Multiple social factors that affect perception of Latina women's speech. Undergraduate Honors Thesis. University of Oregon.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rubin, Donald. 1992. Nonlanguage factors affecting undergraduates' judgements on nonnative English Speaking teaching assistants. *Research in Higher Education* 33(4): 511-531.
- Staum Casasanto, Laura. 2008. Does social information influence sentence processing? 30th Annual Meeting of the Cognitive Science Society 30:799-804.
- Strand, Elizabeth. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18(1):86–100.
- Strand, Elizabeth. 2000. Gender stereotype effects in speech processing. Doctoral Dissertation, Ohio State University.
- Tortora, Christina, Cecelia Cutler, Bill Haddican, Michael Newman, Beatrice Santorini, and C. A. Diertani, in progress. Corpus of New York City English (CUNY-CoNYCE).
- Wong, Amy Wing-mei. 2007. Two vernacular features in the English of four American-born Chinese. In *U. Penn Working Papers in Linguistics 13.2: Selected Papers from NWAV 35*, 217-230.
- Wong, Amy Wing-mei. 2010. New York City English and Second Generation Chinese Americans. *English Today* 26:3–11.
- Wong, Amy Wing-mei. 2012. The lowering of raised-THOUGHT and the low-back distinction in New York City: Evidence from Chinese Americans. In *U. Penn Working Papers in Linguistics 18.2: Selected Papers from NWAV 40*, ed. H. Prichard, 157-166.
- Wong, Amy Wing-mei, and Lauren Hall-lew. 2014. Regional variability and ethnic identity: Chinese Americans in San Francisco and New York City. *Language and Communication* (35):27-42.

Department of Linguistics
 City University of New York, Graduate Center
 New York, NY 10016
 oortiz90@gmail.com
 bill.haddican@gmail.com